



Fast, accurate and automatic ancient nucleosome and methylation maps with epiPALEOMIX

Hanghøj, Kristian Ebbesen; Seguin-Orlando, Andaine; Schubert, Mikkell; Madsen, Tobias; Pedersen, Jakob Skou; Willerslev, Eske; Orlando, Ludovic Antoine Alexandre

Published in:
Molecular Biology and Evolution

DOI:
[10.1093/molbev/msw184](https://doi.org/10.1093/molbev/msw184)

Publication date:
2016

Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY-NC](#)

Citation for published version (APA):
Hanghøj, K. E., Seguin-Orlando, A., Schubert, M., Madsen, T., Pedersen, J. S., Willerslev, E., & Orlando, L. A. (2016). Fast, accurate and automatic ancient nucleosome and methylation maps with epiPALEOMIX. *Molecular Biology and Evolution*, 33(12), 3284-3298. <https://doi.org/10.1093/molbev/msw184>

Fast, Accurate and Automatic Ancient Nucleosome and Methylation Maps with epiPALEOMIX

Kristian Hanghøj,^{1,2} Andaine Seguin-Orlando,^{1,3} Mikkel Schubert,¹ Tobias Madsen,^{4,5} Jakob Skou Pedersen,^{4,5} Eske Willerslev,^{1,6,7} and Ludovic Orlando^{*,1,2}

¹Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

²Laboratoire d'Anthropobiologie Moléculaire et d'Imagerie de Synthèse, Université de Toulouse, University Paul Sabatier, Toulouse, France

³Danish National High-Throughput DNA Sequencing Center, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

⁴Department of Molecular Medicine (MOMA), Aarhus University Hospital, Aarhus, Denmark

⁵Bioinformatics Research Centre (BiRC), Aarhus University, Aarhus, Denmark

⁶Department of Zoology, University of Cambridge, Cambridge, United Kingdom

⁷Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom

*Corresponding author: E-mail: lorlando@snm.ku.dk.

Associate editor: Connie Mulligan

Abstract

The first epigenomes from archaic hominins (AH) and ancient anatomically modern humans (AMH) have recently been characterized, based, however, on a limited number of samples. The extent to which ancient genome-wide epigenetic landscapes can be reconstructed thus remains contentious. Here, we present epiPALEOMIX, an open-source and user-friendly pipeline that exploits post-mortem DNA degradation patterns to reconstruct ancient methylomes and nucleosome maps from shotgun and/or capture-enrichment data. Applying epiPALEOMIX to the sequence data underlying 35 ancient genomes including AMH, AH, equids and aurochs, we investigate the temporal, geographical and preservation range of ancient epigenetic signatures. We first assess the quality of inferred ancient epigenetic signatures within well-characterized genomic regions. We find that tissue-specific methylation signatures can be obtained across a wider range of DNA preparation types than previously thought, including when no particular experimental procedures have been used to remove deaminated cytosines prior to sequencing. We identify a large subset of samples for which DNA associated with nucleosomes is protected from post-mortem degradation, and nucleosome positioning patterns can be reconstructed. Finally, we describe parameters and conditions such as DNA damage levels and sequencing depth that limit the preservation of epigenetic signatures in ancient samples. When such conditions are met, we propose that epigenetic profiles of CTCF binding regions can be used to help data authentication. Our work, including epiPALEOMIX, opens for further investigations of ancient epigenomes through time especially aimed at tracking possible epigenetic changes during major evolutionary, environmental, socioeconomic, and cultural shifts.

Key words: ancient DNA, epigenome, DNA methylation, nucleosome.

Background

In the last decade, ancient DNA (aDNA) research has been truly revolutionized as the information retrieved is no more limited to the sole mitochondrial DNA (mtDNA) or few nuclear fragments carrying SNPs of interest. Complete genomes and up to several millions of SNPs can now be characterized instead (see, Orlando et al. 2015, for a review). This is because our ability to retrieve traces of genetic material has tremendously improved following the development of High-Throughput DNA Sequencing (HTS) technologies (Metzker 2010) and experimental methods tailored to the recovery and manipulation of ultrashort and damaged DNA molecules (Meyer et al. 2012; Dabney et al. 2013; Gansauge and Meyer 2013, 2014; Gamba et al. 2016). As a result, the time barrier for

genome-scale analyses has been pushed backwards into the Middle Pleistocene (Dabney et al. 2013; Orlando et al. 2013; Meyer et al. 2014, 2016) and samples that were previously not amenable to DNA analyses are becoming so (Hofmanová et al. 2016; Lazaridis et al. 2016).

In addition to recovering genome-scale genetic information from ancient individuals, it has also become possible to analyze their epigenetic profiles (see, Orlando et al. 2015; Gokhman et al. 2016, for reviews). The first analysis of cytosine methylation patterns at the single-base resolution was performed on a ~26,000-year-old bison bone using bisulfite allelic sequencing (Llamas et al. 2012). Bisulfite conversion of unmethylated cytosines into uracils, however, is not perfectly suited to aDNA material, which is generally degraded and available in very limited amounts. Minimum DNA

concentration requirements, thus, apply before consistent measures of cytosine methylation can be recovered with this technology (Smith et al. 2015). Other direct approaches, such as enrichment procedures based on methylated binding domains (MBD), have been used (Seguin-Orlando, Gamba, et al. 2015) but their performance is generally limited given the extensive fragmentation of aDNA templates. In contrast, indirect approaches exploiting the chemical degradation reactions affecting DNA post-mortem are particularly suited to the chemical nature of aDNA. Such methods track in silico the signatures of nucleosome protection and the natural conversion of unmethylated cytosines into uracils that take place after death. They recently delivered the first genome-wide methylation and nucleosome maps (Gokhman et al. 2014; Pedersen et al. 2014), providing the first insights at how the regulation of gene expression changed in evolutionary times.

The evidence supporting such genome-wide methylation and nucleosome maps is, however, limited to only three individuals. These include a 4-kyr-old Palaeo-Eskimo (Pedersen et al. 2014), a ~50-kyr-old Neanderthal and a ~40-kyr-old Denisovan (Gokhman et al. 2014). Analyses of a larger number of specimens, spanning a wider temporal range, tissue types and preservation conditions, are thus essential to establish the true potential (and limitations) of ancient epigenomics.

Albeit limited, previous work has suggested that gene expression levels and the age at death of ancient individuals might be inferred based on ancient epigenetic signatures (Pedersen et al. 2014). Additionally, potentially key DNA methylation changes for the differentiation of closely related lineages have started to be unveiled (Gokhman et al. 2014). These pioneering studies, and the growing evidence supporting epigenetic differentiation between major human population groups (Heyn et al. 2013; Fagny et al. 2015), open for wider research on how epigenetic signatures have changed on a temporal scale in the face of important environmental and cultural changes, such the Neolithic transition and/or responses to major disease outbreaks (Orlando and Willerslev 2014).

With exponentially increasing sequence data sets from ancient individuals (reviewed in Orlando et al. 2015), an automated tool is required to facilitate and standardize the analysis of ancient epigenomes. RoAM (Reconstruction of Archaic Methylation; <http://carmelab.huji.ac.il/software/RoAM/roam.html>; last accessed September 1, 2016) provides the first of such tools, allowing the reconstruction of genome-wide ancient methylation maps, but its current implementation is restricted to archaic hominins (AH). In this article, we present epiPALEOMIX, an open-source and user-friendly software package that automates the characterization of ancient epigenomes. The epiPALEOMIX pipeline implements all the methods used to characterize the first ancient epigenome, that from the hair of a 4-kyr-old Palaeo-Eskimo (Pedersen et al. 2014), including methylation mapping, nucleosome calling and phasogram analyses. It can be readily applied to any BAM alignment file obtained from ancient specimens. We apply epiPALEOMIX to the shotgun sequence data underlying a large subset of ancient genomes characterized with an

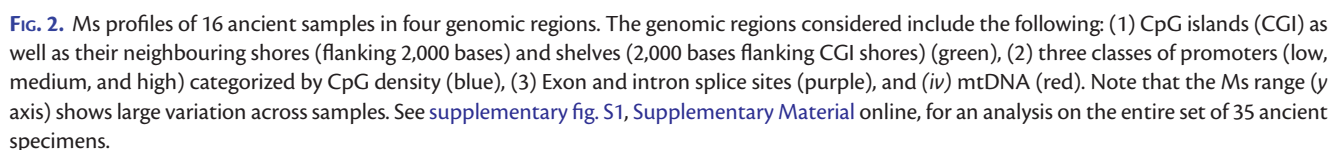
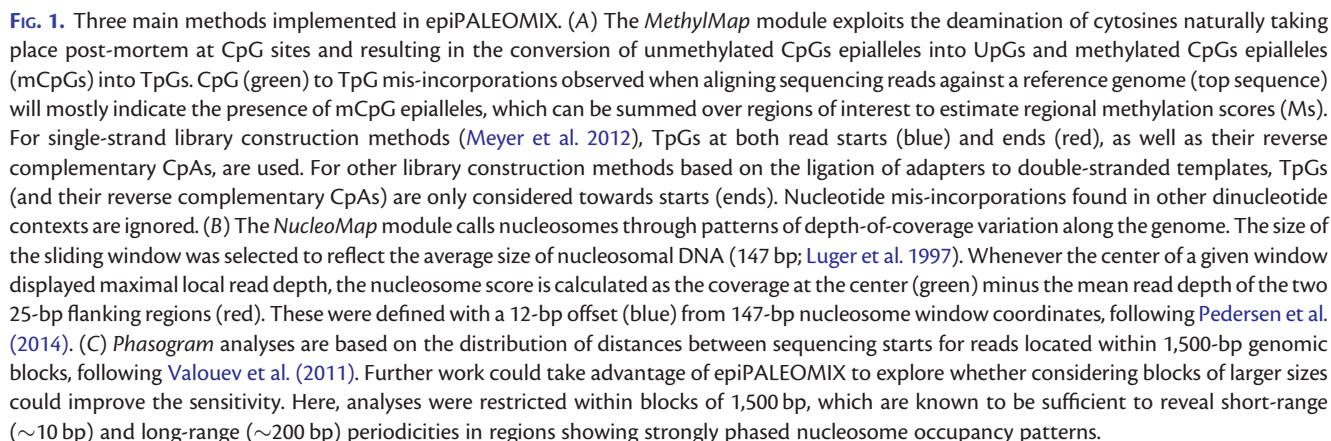
average depth-of-coverage >2-fold (2.19- to 42.0-fold). Our study thus represents the largest ancient epigenomic study, including 35 samples from AH, ancient anatomically modern humans (AMH), horses and aurochs. Compared with previous analyses, the samples considered here significantly increase the temporal resolution (from 100 years to over 50 kyr) and geographical (Africa, North America, Central Asia, Europe, Greenland, and Siberia) range of epigenomic profiling.

New Approaches

In this study, we present epiPALEOMIX, a new open-source package for the analysis of ancient epigenomes, which encloses one optional and three main modules. The *MethylMap* module (fig. 1A) implements the methodology from Pedersen et al. (2014) and Gokhman et al. (2014) to calculate methylation scores (Ms) and reconstruct methylation profiles within regions provided by users as BED coordinate files. Such files can correspond to genomic regions of interest, such as promoters, gene bodies, CpG Islands (CGIs) (fig. 2), but also to sliding genomic windows (fig. 10). The methodology exploits the relative number of CpG→TpG read mis-incorporations found within a given region, which is assumed to mostly reflect post-mortem deamination events at methylated epialleles (Seguin-Orlando, Hoover, et al. 2015). The *MethylMap* module can accommodate single-stranded and double-stranded DNA library types and can also restrict the analyses to specific read positions, which proves useful in cases where particular read positions show inflated error rates. These features, and the direct applicability of the methodology to nonarchaic hominin sequence data, represent important increments over RoAM. As specifically developed for the analysis of epigenomic signatures from aDNA data, epiPALEOMIX also fills a niche that is presently not covered through the multiple Bioconductor tools tailored to the detection of methylation and nucleosome regions (Gentleman et al. 2004; Huber et al. 2015).

In addition to *MethylMap*, epiPALEOMIX provides two modules to reconstruct patterns of nucleosome positioning along the genome. The first module, called *NucleoMap*, implements the methodology from Pedersen et al. (2014) to map peaks of nucleosome occupancy (fig. 1B). It builds on the natural DNA decay taking place post-mortem, resulting in an over-representation of nucleosome-protected DNA regions post-sequencing (Dong et al. 1997; Pedersen et al. 2014). More specifically, *NucleoMap* exploits patterns of coverage variation along the genome, where nucleosomes are inferred at positions maximizing depth variation within 147 nucleotide blocks, reflecting the size of a nucleosome dyad, and their flanking regions. The analysis can be performed genome-wide or within pre-defined genomic regions (as shown on fig. 7). The second module implements the phasogram analyses from Valouev et al. (2011) and aims at detecting periodicity patterns between DNA fragments and from adjacent nucleosomes (fig. 1C).

We also provide an optional module, *GCcorrect*, that builds on previous methodology (Benjamini and Speed 2012) to



The epiPALEOMIX package is written in Python 2.7 and builds on the implementation of the PALEOMIX pipeline (Schubert et al. 2014). It supports multi-threading and requires a simple makefile as input indicating the location of sequence

alignment files (BAM format) and the parameters of the analyses to be run. The epiPALEOMIX package is compatible with the analysis of HTS data generated from shotgun and capture-enrichment experiments. Analyses restricted to a single or a minimum number of loci are typically performed within minutes while up to half an hour is necessary to investigate approximately half a million loci scattered along the genome ([supplementary table S1, Supplementary Material online](#)).

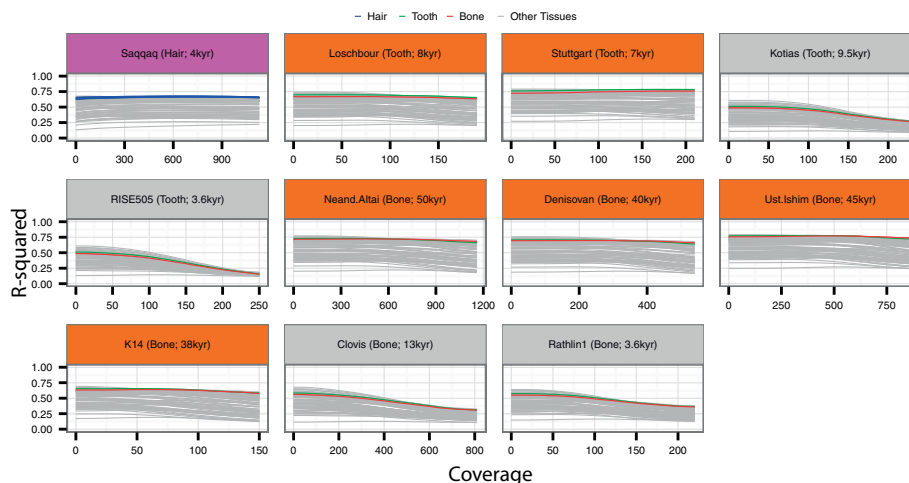


Fig. 3. Tissue specific methylation profiles correlate with Ms. The Rsquared correlation coefficients between Ms scores and a wide range of methylation profiles from fresh tissues are provided across a range of coverage thresholds, representing the minimal number of CpG and TpG sites found within the proximal 15 positions of reads in a considered region. RRBS tooth and bone profiles are shown in green and red, respectively, and 450K Illumina array hair profiles are shown in blue. See [supplementary fig. S3, Supplementary Material](#) online, for an analysis on the entire set of 30 ancient AMH and AH.

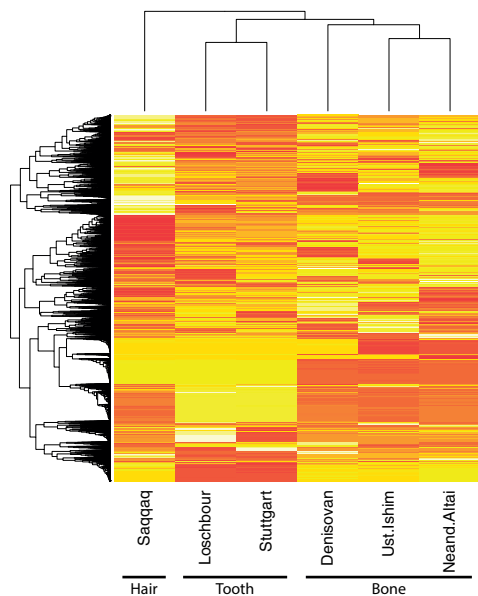


Fig. 4. Unsupervised hierarchical clustering based on Rs ratios. Rs ratios represent the ratio between Ms scores of gene bodies and promoter regions. The tissues sampled are shown below sample names.

The epiPALEOMIX package is available at <https://bitbucket.org/khanghoj/epipaleomix> (last accessed September 1, 2016), together with an extensive companion manual, running examples, BED files, and mappability maps to ensure reproducibility of all analyses presented in this study.

Results and Discussion

Preliminary Notes

In the following sections, we applied epiPALEOMIX to HTS data generated following three main experimental procedures to explore the presence of genome-wide epigenetic signatures

in a wide range of ancient samples, including different tissue types (hair, bones and teeth), geographic origins and DNA preservation conditions. In the first experimental procedure, aDNA extracts were treated with USER (New England Biolabs), a commercial DNA repair mix containing Uracil DNA Glycosylase (UNG) and Endonuclease VIII (EndoVIII) (method 1-USER, orange figure header backgrounds). This treatment breaks DNA strands 3' of uracil residues and eliminates most C→T nucleotide mis-incorporations post-sequencing (Briggs et al. 2010; Gokhman et al. 2014), except at template termini (Meyer et al. 2012) and methylated cytosines. The next two approaches do not involve USER-treatment prior to library construction. In method 2-Regular (gray figure header backgrounds), aDNA libraries are amplified using DNA polymerases that can bypass uracil residues, such as AmpliTaq Gold (Thermo Fisher). In method 3-*Phusion* (purple figure header backgrounds), aDNA libraries are amplified using DNA polymerases that cannot bypass uracil residues, such as *Phusion*. Both methods 1 and 3 have been shown to provide a way to estimate regional methylation levels through Ms (Pedersen et al. 2014; Seguin-Orlando, Hoover, et al. 2015).

Our sample set includes 28 ancient human samples and 2 AH (table 1). We also included five nonhuman organisms, including four ancient equine specimens (~0.1–42 kya; Schubert et al. 2014; Der Sarkissian et al. 2015; Librado et al. 2015) and a single aurochs (~6.7 kyr old) (Park et al. 2015).

Regional Methylation Patterns

We stratified the human genome into four main genomic classes expected to show strong differential methylation levels but low cross-tissue variation (Ziller et al. 2013). We first investigated CGIs, including their up- and downstream flanking shores (0–2 kb) and shelves (2–4 kb). CGIs are well-known hypomethylated genomic regions, showing gradually increasing methylation levels from shores to shelves (Deaton and Bird 2011). We next focused on three categories of

Table 1. Hominin Samples Used in This Study.

Time period/culture	Type	Number	References
Iron age (1.4 kya)	AMH	1	Allentoft et al. (2015)
Bronze age (3–4.7 kya)	AMH	15	Gamba et al. (2014) Allentoft et al. (2015) Günther et al. (2015) Cassidy et al. (2016)
Early Neolithic (7–7.25 kya)	AMH	2	Gamba et al. (2014) Lazaridis et al. (2014)
Mesolithic (7–9.5 kya)	AMH	4	Lazaridis et al. (2014) Olalde et al. (2014)
Upper Paleolithic (13.5–45 kya)	AMH	3	Jones et al. (2015) Fu et al. (2014) Seguin-Orlando et al. (2014)
Saqqaq (4 kya)	AMH	1	Jones et al. (2015) Rasmussen et al. (2010)
Clovis (12 kya)	AMH	1	Rasmussen et al. (2014)
pre-Bantu African (4.5 kya)	AMH	1	Gallego Llorente et al. (2015)
Upper Paleolithic (~40 kya)	AH (Denisovan)	1	Meyer et al. (2012)
Upper Paleolithic (~50 kya)	AH (Neanderthal)	1	Prüfer et al. (2014)

NOTE.—AMH, anatomically modern human; AH, archaic hominin.

promoter regions showing *High*, *Intermediate* and *Low* CpG density and %GC content, which are known to be inversely correlated with methylation levels (Ball et al. 2009). We then contrasted Ms in exons, introns and their splice sites, as 5' (3') exonic DNA shows higher methylation levels than 5' (3') intronic DNA (Laurent et al. 2010). Finally, we examined mtDNA as this locus shows minimal, if any, methylation (Rebelo et al. 2009).

Almost all ancient HTS data sets investigated displayed the expected methylation patterns, regardless of the molecular tools used for generating data (fig. 2 and supplementary fig. S1, Supplementary Material online). This was in line with previous work for all specimens prepared following USER treatment (method 1-USER) and *Phusion* amplification (method 3-*Phusion*) (Gokhman et al. 2014; Pedersen et al. 2014). However, this had never been observed in the absence of such experimental procedures (method 2-Regular), as CpG→TpG mis-incorporations are introduced at both methylated and unmethylated CpGs. That expected regional methylation levels are found with this method corroborates experimental evidence (Dianov et al. 1994; Seguin-Orlando, Gamba, et al. 2015) that post-mortem cytosine deamination rates are significantly faster in methylated CpG contexts (relative to nonmethylated CpG contexts). As a result, samples such as the CGG10397 horse specimen that lived in the 19th century AD show expected methylation profiles at CGIs, CGI shores and shelves and promoters, despite its relative young age (fig. 2).

We note that in method 1-USER and method 3-*Phusion* data, mtDNA Ms levels are, as expected, reduced in comparison to those observed in hypomethylated regions of the nuclear genome. However, for all samples prepared with method 2-Regular where mitochondrial sequencing data were available, such mtDNA Ms levels were found comparable, and even greater, to those observed in the nuclear genome. This is in agreement with the expected accumulation of C→T mis-incorporations at nonmethylated sites, and

recent analyses showing higher deamination rates in mtDNA than nuclear DNA (Kistler et al. 2015).

Finally, we repeated the same analyses on MBD-capture sequence data that we generated from USER-treated DNA extracts of the Palaeo-Eskimo Saqqaq individual. Even though only 622,650 high-quality reads were found to map uniquely against the human genome, we recovered expected methylation signatures at CGIs, promoters and mtDNA (supplementary fig. S2, Supplementary Material online). This demonstrates the applicability of epiPALEOMIX to capture data.

Tissue-Specific Methylation Profiles

We next used epiPALEOMIX to investigate whether the sequence data underlying the 30 ancient AMH and AH individual genomes (table 1) showed tissue-specific methylation profiles. To achieve this, we compared Ms to 101 Reduced Representation Bisulfite Sequencing (RRBS) profiles obtained from 45 tissues and cell types (Meissner et al. 2008). We also included 61 450K Illumina array profiles of 22 tissues (Sandoval et al. 2011; Slieker et al. 2013) to also represent hair methylomes, which were absent from the RRBS data. Ms values were calculated within 2-kb windows centered at each position present on the RRBS and/or 450K Illumina array data sets (Pedersen et al. 2014). We then built linear models to identify the modern tissues showing methylation profiles most similar to ancient samples (fig. 3 and supplementary fig. S3, Supplementary Material online).

Ms in the high-coverage Palaeo-Eskimo hair sample showed maximal correlation (R -squared = 0.65–0.66) against the five hair tissues (fig. 3), in line with earlier analyses (Pedersen et al. 2014). Modern methylation profiles of osteoblasts showed the strongest correlation with the ~45-kyr-old Ust.Ishim individual out of the 101 methylomes tested (R -squared = 0.74–0.76). This profile also showed the second highest correlation observed in both the Neanderthal (R -squared = 0.69–0.72) and Denisovan (R -squared = 0.66–0.69), in line with earlier analyses (Gokhman et al. 2014).

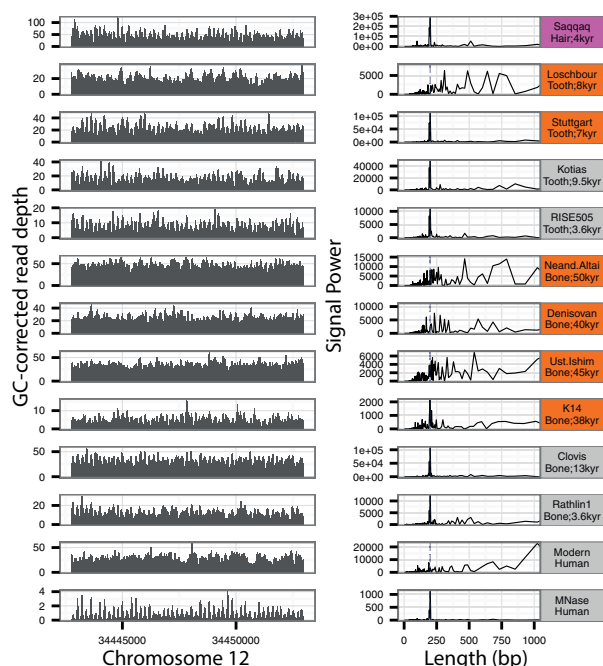


Fig. 5. Nucleosome occupancy patterns at a well-phased nucleosome array. (Left) GC-corrected read depth profiles across a sub-region of an alpha-satellite repeat region encompassing positions 34,439,733–34,559,733 on chromosome 12 (hg19). (Right) power spectral density plot across uniquely mappable regions of the nucleosome array. A vertical dashed line is shown at 200 bp to facilitate readability. Note that the periodicity power (y axis) shows large variation across samples. See [supplementary fig. S5, Supplementary Material online](#), for an analysis on the entire set of 30 ancient AMH and AH.

(fig. 3 and [supplementary table S2, Supplementary Material online](#)). Interestingly, the correlation with the osteoblast profile was still strong for the Upper Paleolithic K14 individual (R -squared = 0.58–0.63; 5th highest R -squared, [supplementary table S2, Supplementary Material online](#)), although sequenced to only $2.42\times$ depth-of-coverage. The same was true for all bone samples prepared with method 2-Regular (0.22–0.47), except ATP2, which showed distinctively lower R -squared values across all modern profiles investigated (fig. 3 and [supplementary table S2](#) and fig. S3, [Supplementary Material online](#)). This is in accordance with the limited difference in M_s found for this sample within expected hypo- and hyper-methylated regions ([supplementary fig. S1, Supplementary Material online](#)), and provides an example where ancient methylation signatures are not fully exploitable.

USER-treated Stuttgart and Loschbour teeth samples showed maximal correlation with a methylation profile retrieved from modern gingival, a tissue naturally in contact with teeth (R -squared = 0.76–0.78 and 0.65–0.70, respectively) and second highest for the tooth Motala sample (R -squared = 0.43–0.57). We also found strong correlation levels between the ~ 9.5 -kyr-old Kotias tooth methylation profile and that from modern gingival tissues (R -squared = 0.27–0.51). This contrasts with the ~ 7.5 -kyr-old LaBranca sample and a series of Bronze Age teeth (RISE395, RISE493, RISE495,

RISE496, RISE497, RISE505, RISE511 and RISE523), which were not USER-treated and showed relatively low R -squared values at maximal coverage (≤ 0.20) across all modern profiles investigated ([supplementary fig. S3, Supplementary Material online](#)).

Altogether, we found that for aDNA data prepared with methods 1-USER and 3-Phusion, ancient methylation profiles were reminiscent of the modern methylomes of similar tissues or cell types for the majority of the samples investigated. HTS data generated with method 2-Regular showed lower correlation levels, which also significantly dropped with increasing coverage thresholds. We thus do not recommend this method for investigating ancient methylomes, especially for samples similar to ATP2 and RISE395, both sequenced to an average 4.08- and 3.54-fold coverage-of-depth, even though they show relatively high expected numbers of cytosine deaminations at overhangs (0.70–1.53; see below). Interestingly, we noticed that tooth gingival and bone RRBS methylation profiles were the most correlated (0.95, [supplementary fig. S4, Supplementary Material online](#)) but also showed high correlation levels (>0.90) with samples from other tissues types, such as blood, brain and muscles. In the future, this will help researchers test whether the observed methylation changes were tissue-specific or also likely affected other tissues. This may also provide opportunities to infer ancient methylation profiles for soft tissues, despite the large majority of ancient subfossils consisting of calcified material.

To investigate whether methylation levels could reflect different gene expression profiles across tissues, we considered 19,270 human gene models and calculated R_s ratios as proposed by [Pedersen et al. \(2014\)](#). These correspond to ratios of M_s between gene bodies and promoters, and are known to correlate with gene expression in human cell cultures, with housekeeping genes showing maximal R_s ratios ([Ball et al. 2009](#)). Using R_s ratios, samples sequenced at high average depth-of-coverage ($>19\times$) and prepared with methods 1-USER and 3-Phusion were found to cluster according to the fossil material extracted, with three main groups represented by hair (Palaeo-Eskimo Saqqaq), teeth (Loschbour and Stuttgart) and bones (Neand. Altai, Denisovan and Ust. Ishim) (fig. 4).

Interestingly, the top 95th quantile of R_s ratios for the three ancient bone methylomes included genes encoding collagen proteins (I, II, III, V, and XI), suggesting constitutive expression levels in agreement with their structural role in the extracellular matrix of bones and other connective tissues ([Niyibizi and Eyre 1994](#)). These genes were not part of the top 95th quantile in the hair Palaeo-Eskimo individual, which instead returned *Keratin 17*, *32*, *33a*, *Periplakin*, and *Plectin*, all representing proteins related to hair and hair follicle, in line with findings by [Pedersen et al. \(2014\)](#). Unfortunately, we could not investigate genes highly expressed in tooth enamel (*AMELX*, *TUFT1*, and *AMBN*) ([Delsuc et al. 2015](#)) and tooth dentine (*Oc* and *Dspp*) ([Hoffmann et al. 2001](#)), as they did not pass our filtering thresholds. Altogether, these analyses plead in favor of the tissue-specificity of R_s scores.

However, when investigating whether changes in either bone (Neand. Altai, Denisovan and Ust. Ishim) or tooth

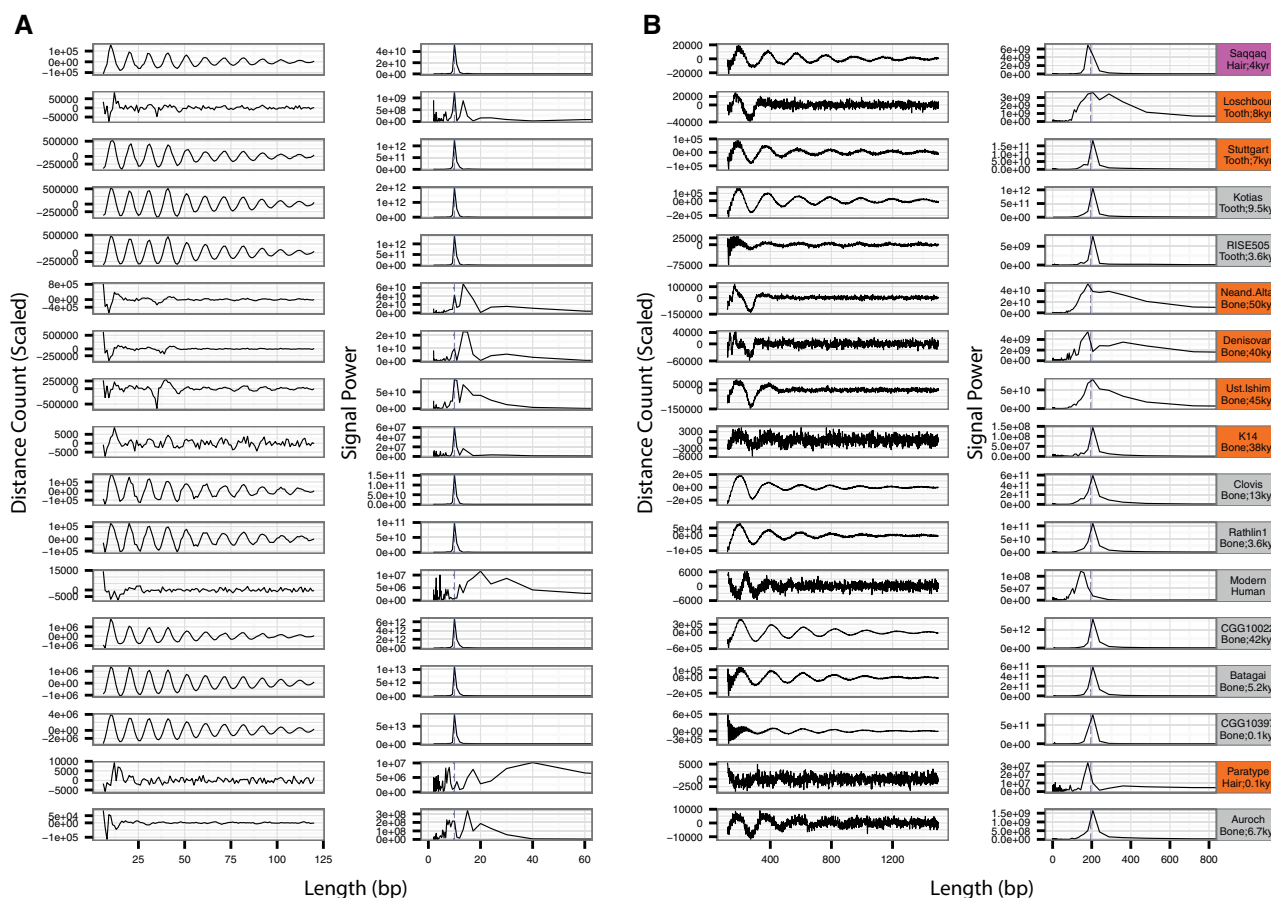


Fig. 6. Phasogram analysis of gene bodies. (A) Short-range phasograms (Left) and their respective spectral density plots (Right). Vertical dashed line at 10 bp. (B) Long-range phasograms (Left) and their respective spectral density plots (Right). A vertical dashed line is shown at 200 bp to facilitate readability. Note that the periodicity power (y axis of the right panels) shows large variation across samples. See [supplementary fig. S7, Supplementary Material](#) online, for an analysis on the entire set of 35 ancient specimens.

(Loschbour and Stuttgart) expression levels could be detected through time, we only found genes involved in high-level biological functions (data not shown). We believe that this reflects the combination of gene expression reprogramming for ubiquitous proteins common to a range of tissues as well as the presence of limited coverage at gene bodies for a majority of the genes considered. Further work is, thus, needed before ancient proxies of gene expression, such as Rs ratios, could be interpreted biologically.

Nucleosome Occupancy and Phasing

We first investigated whether patterns of read depth variation reflected nucleosome occupancy, focusing on an alpha-satellite repeat region on the human chromosome 12 showing well-phased nucleosome arrays across a whole range of tissues after Micrococcal Nuclease (MNase) treatment (Gaffney et al. 2012). We found that for a majority of ancient hominin samples (25/30, representing all samples but Motala, Loschbour, Neandertal, Denisovan and Ust-Ishim), GC-corrected depth-of-coverage profiles were highly correlated to lymphoblastoid cell MNase profiles (Valouev et al. 2011; Gaffney et al. 2012) (Pearson correlation coefficient, $PCC = 0.29-0.68$, P values $< 1e-16$) (fig. 5 and [supplemen-](#)

[tary figs. S5 and S6, Supplementary Material](#) online). These samples, except RISE98, showed a predominant distance of ~ 200 bp between consecutive peaks of read depth, in line with the MNase treated profile and the length of nucleosome and linker regions (representing 147 and ~ 50 bp, respectively) (fig. 5, right column and [supplementary fig. S5, right column, Supplementary Material](#) online). Importantly, we found no or negative correlation when comparing the ancient samples to the sequence data of a modern human individual from the 1,000 human genome project that we used as negative control ($PCC = -0.26-0.08$, P values $< 1e-5$).

Secondly, we used epiPALEOMIX to carry out a phasogram analysis on our entire panel of samples, including the four equids and the aurochs. Phasograms display the distribution of distances between read start positions on the same strand (Valouev et al. 2011) and can help detect fragmentation periodicities. In regions where nucleosomes are well phased, such as gene bodies (Jiang and Pugh 2009), one nucleotide residue is expected to be more exposed to hydrolysis for every DNA helix turn (Wang et al. 2008; Jiang and Pugh 2009; Pedersen et al. 2014; Orlando et al. 2015), leading to short-range (~ 10 bp) periodicity patterns in ancient HTS data (Pedersen et al. 2014). Strong nucleosome phasing is also

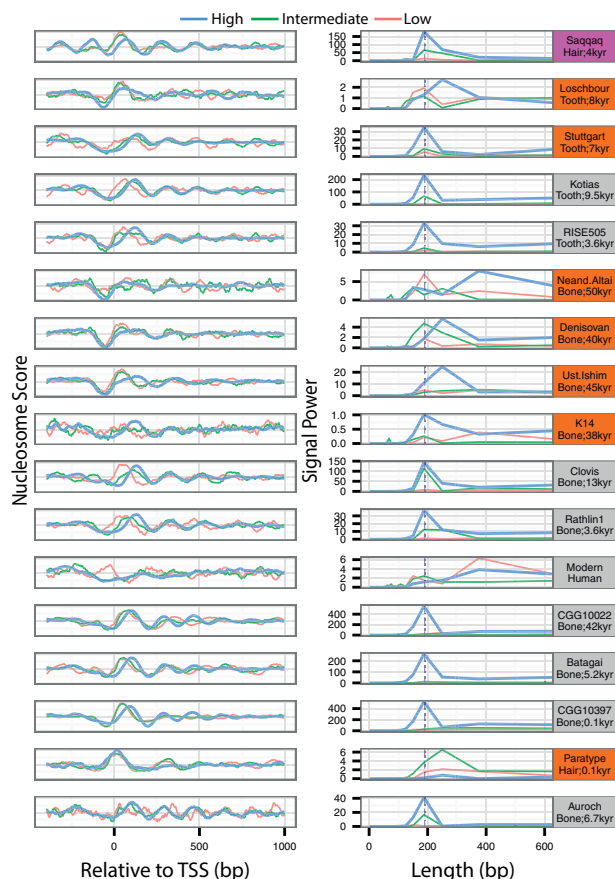


Fig. 7. Sliding nucleosome score of GC-corrected read-depth variation around TSSs. (Left) Promoters were stratified following Ball et al. (2009) in three classes according to their CpG density and %GC content (Low: green, Intermediate: red, High: blue). (Right) Power spectral density plot from TSS and 1-kb downstream. A vertical dashed line is shown at 190 bp to facilitate readability. See [supplementary fig. S8, Supplementary Material](#) online, for an analysis on the entire set of 35 ancient specimens.

expected to introduce long-range periodicity patterns reflecting the distance between consecutive nucleosomes, on average every ~ 200 bp (Valouev et al. 2011). We found short-range periodicity patterns for the same set of individual genomes displaying ~ 200 bp periodicity in the alpha-satellite repeat region used above (fig. 5 and [supplementary fig. S5, Supplementary Material](#) online), with the addition of Motala, Loschbour, Ust.Ishim, and three horses (CGG10022, Batagai, CGG10397), summing up to 30/35 samples. Most of these samples, excepting Loschbour, Motala, RISE511 and RISE495, but with the addition of the horse Paratype and Aurochs also showed strong long-range periodicity patterns (totaling 27/35 samples). Neand.Altai, Denisovan and Ust.Ishim showed more noisy ~ 200 bp periodicity signals. Of note, our modern negative control showed no short-range periodicity (fig. 6 and [supplementary fig. S7, Supplementary Material](#) online) but a ~ 144 -bp long-range density peak, in line with phasogram analyses on modern data (Valouev et al. 2011).

Thirdly, we investigated nucleosome occupancy patterns around TSS, where strong nucleosome phasing is found in *Saccharomyces cerevisiae* house-keeping genes (Jiang and

Pugh 2009) and to some extent, in humans (Ozsolak et al. 2007). To achieve this, we called nucleosomes and looked for long-range periodicity patterns within the three promoter categories described above (fig. 7 and [supplementary fig. S8, Supplementary Material](#) online). We found consistent ~ 190 -bp periodicities from TSS for the *High* promoter category across the same set of samples that showed long-range periodicity signals in previous phasogram analyses, with the single exception of the horse Paratype, which is amongst the most recent samples (fig. 7 and [supplementary fig. S8, Supplementary Material](#) online). For the remaining two promoter categories (*Low* and *Intermediate*), we find 3- to 200-fold weaker signal powers of Fourier periodicity, in line with their known, less pronounced, nucleosome occupancy and phasing patterns in living cells (Hesson et al. 2014).

We next followed Snyder et al. (2016) and binned human TSS in five increasing expression quantiles. For 21 out of the 30 hominin samples investigated, we found a much stronger periodicity in nucleosome positioning for high expression categories relative to low expression categories. Therefore, a majority of the ancient samples support the known relationship between nucleosome occupancy and phasing around TSS and gene expression ([supplementary fig. S9, Supplementary Material](#) online).

Finally, we note that we systematically failed retrieving the expected signatures of nucleosome protection from the Neand.Altai, Denisovan, Loschbour, Paratype, Motala and RISE98 sequence data. Given that (1) these samples encompass the full temporal range investigated here, and (2) short-range periodicities have been described in cell-free DNA circulating in the blood of living individuals (Snyder et al. 2016), the presence of such signatures therefore does not appear to be age-dependent but rather damage dependent.

NucleoMap versus WPS

The Windowed Protection Score (WPS) method has been recently developed to call nucleosomes in DNA fragments circulating freely in the blood of living individuals (Snyder et al. 2016). Cell-free DNA fragments generally show strong patterns of nucleosome protection and a short-range periodicity in phasogram analyses. We thus investigated whether the WPS method could complement epiPALEOMIX in characterizing ancient nucleosome maps. We restricted our analyses to the well-phased nucleosome array found within the alpha-satellite repeat region on chromosome 12, running WPS on a wide range of possible window protection sizes (10, 20, 30, 50, 80, 120 bp). Firstly, we found that, regardless of the sizes considered, the WPS method called fewer peaks (50–480) than the 586 nucleosomes found in the conserved nucleosome array following MNase treatment. The *NucleoMap* method from epiPALEOMIX consistently called a larger number of peaks, except for the Saqqaq sample for which similar nucleosome counts were found in both epiPALEOMIX and WPS (protection windows 30–80 bp; [supplementary fig. S10, Supplementary Material](#) online). Secondly, nucleosome dyads from the MNase treated sample were found to be physically closer to those called with epiPALEOMIX than those called with WPS ([supplementary fig. S11, Supplementary Material](#)

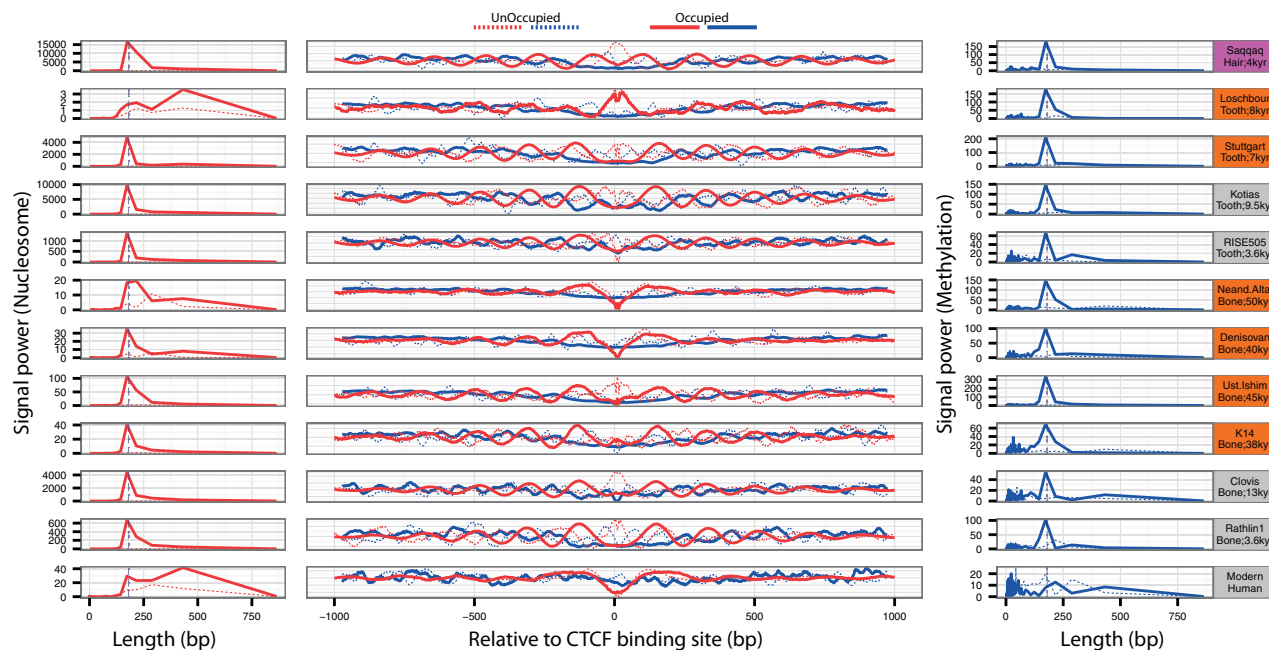


FIG. 8. Methylation profiles and GC-corrected read-depth variations at CTCF regions (61 kb). (Left) Spectral density plot for nucleosomes (red). (Central) Methylation (blue) and nucleosome occupancy (red) at CTCF regions. (Right) Spectral density plot for Ms scores (blue). The profiles recovered at occupied (unoccupied) CTCF-binding sites (Fu et al. 2008) are reported in solid (dashed) lines. A vertical dashed line is shown at 180 bp to facilitate readability. Note that the periodicity power of nucleosome occupancy patterns and methylation levels show large variation across samples. See [supplementary figs. S12 and S13, Supplementary Material](#) online, for an analysis on the entire set of 30 ancient AMH and AH.

online). This was true across all protection windows. Notably, WPS nucleosome calls with the smallest protection windows (≤ 30 bp) were closest to epiPALEOMIX and MNase called peaks, which is dramatically smaller than the 120 bp or higher used to identify nucleosome positioning in cell-free DNA (Snyder et al. 2016). Therefore, our results suggest that *NucleoMap* should be preferred to the WPS method when analyzing aDNA data, probably due to the implicit assumption in WPS that no secondary fragmentation takes place after initial fragmentation by apoptosis.

Ancient Methylation and Nucleosome Profiles at CTCF Binding Sites

In the following, we focused on the 2-kb regions surrounding CTCF binding sites, as such regions are known to exhibit strong and inversely correlated methylation and nucleosome patterns whenever bound to the CTCF protein (Fu et al. 2008; Kelly et al. 2012). We limited our analyses to ancient AMH and AH individuals in the absence of known coordinates for occupied and un-occupied CTCF regions in horse and cattle. No particular nucleosome patterns were found at unoccupied CTCF regions (fig. 8, dashed lines and [supplementary fig. S12, Supplementary Material](#) online), suggesting that DNA fragmentation is largely random in the absence of nucleosome protection. This contrasts with the strong ~ 180 bp and out-of-phase periodicities found for both methylation and nucleosome occupancy within occupied CTCF regions (fig. 8, solid lines and [supplementary fig. S13, Supplementary Material](#) online). Two samples (Loschbour, RISE98) stand as exceptions, as no clear ~ 180 -bp periodicity in nucleosome occupancies were found, in line with the analyses above

showing no obvious nucleosome protection patterns for these samples. Four of the Bronze Age individuals prepared with Method 2-Regular showed the expected pattern of nucleosome occupancy but lacked the methylation signature (RISE523, RISE511, RISE495, RISE395; [supplementary fig. S13, Supplementary Material](#) online, blue). These were thus excluded from the following analyses.

Using a downsampling procedure, we found that a minimal number of 200,000–1,600,000 reads on occupied CTCF regions are generally sufficient to recover the expected out-of-phase periodicity signature ([supplementary table S3, Supplementary Material](#) online). This minimal number of reads appears to be inversely correlated with the expected number of deamination-driven nucleotide mis-incorporations found at overhangs (fig. 9, method 1-USER and method 3-Phusion, *R*-squared correlation coefficients = 0.86, *P* value = 0.0016; method 2-Regular, *R*-squared correlation coefficients = 0.35, *P* value = 0.0070). This suggests post-mortem degradation levels as the main driving factor, and that rather than absolute minimal read number requirements, degradation parameters should be considered when checking for such patterns at occupied CTCF regions. We recommend that these patterns can be used as a further authentication criterion for aDNA when a sufficient number of deamination-driven nucleotide mis-incorporations are found at overhangs. Pending further data, we note that no out-of-phase profiles could be recovered for RISE511, RISE495, RISE497, and RISE98, which all show low mis-incorporation rates at overhangs (0.39–0.49) ([supplementary table S3, Supplementary Material](#) online). We thus recommend the sequencing depth reported above and a minimum number

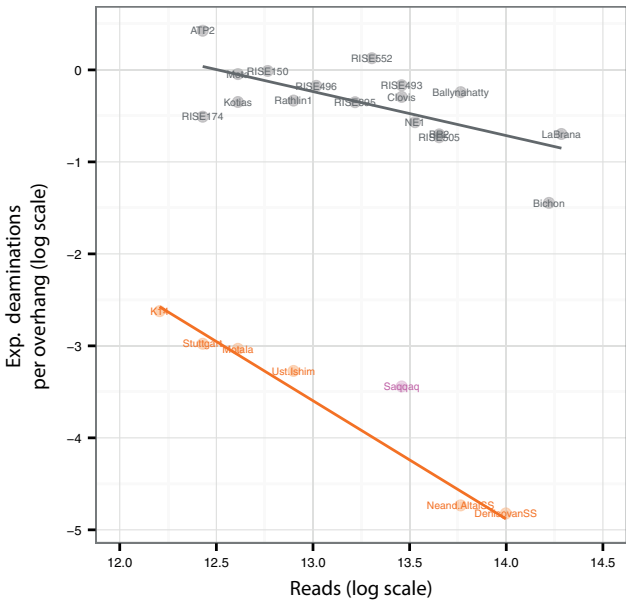


Fig. 9. Detection of epigenetic signatures at occupied CTCF-binding sites depends on post-degradation levels. Individual genomes prepared following method 2-Regular are shown in gray, in contrast to those prepared following methods 1-USER and method 3-Phusion, which are shown in brown and purple, respectively. The expected number of mis-incorporations per overhang region is calculated using the formula provide in the “Methods” section (Regional methylation levels). The x-axis is provided in a natural log scale.

of ~ 0.48 mis-incorporations per overhang should be required for samples prepared with method 2-Regular. That the expected out-of-phase signal between nucleosome occupancy patterns and methylation levels was detectable for Bichon, albeit this sample only showed 0.24 mis-incorporations per overhang, is not incompatible with the recommended threshold as the sequencing data were trimmed for the two starting and ending read positions, where cytosine deamination rates are maximal (Jones et al. 2015). We should, however, caution that CTCF epigenetic patterns should not substitute the other authentication proxies currently used, for example, based on other features of nucleotide mis-incorporation and DNA fragmentation (Briggs et al. 2007; Krause et al. 2010) and/or direct estimates of contamination levels (Green et al. 2010; Rasmussen et al. 2010; Fu et al. 2013; Renaud et al. 2015: 20; Racimo et al. 2016). In contrast to the latter, but similar to the former, CTCF methylation and nucleosome occupancy patterns can indeed be expected in the case of a mixture of modern DNA contaminants and aDNA templates, as long as a sufficient number of deamination-driven nucleotide mis-incorporations are found at overhangs.

Exploiting Ancient Epigenetic Profiles

Previous work has exploited ancient methylation signatures to (1) infer the age at death of ancient individuals (Pedersen et al. 2014) and (2) identify regions differentially methylated

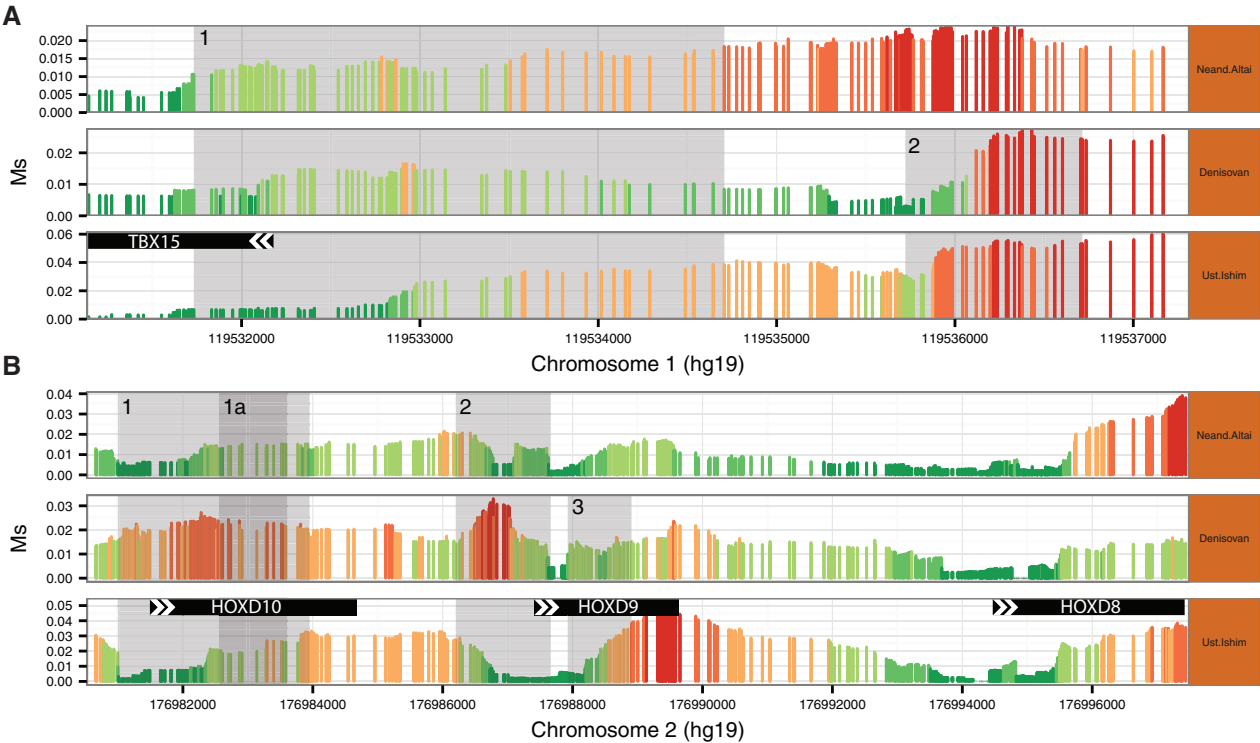


Fig. 10. Sliding window methylation profiles at *TBX15* and the *HOXD* cluster in AH individuals and the 45-kyr-old AMH from Ust Ishim. (A) *TBX15* Ms methylation profiles. (B) Ms methylation profiles at the *HOXD* cluster. Differentially methylated regions between AHs and osteoblast lines from modern AH individuals, as identified by (Gokhman et al. 2014), are highlighted in gray. The DMR1 region from the *HOXD10* gene corresponds to the entire coordinates provided by (Gokhman et al. 2014) in their supplementary information, whereas the DMR1a region corresponds to the same region as highlighted in their main figure.

between AH and modern human individuals (Gokhman et al. 2014). In the next section, we implemented such analyses on our larger data set.

Firstly, we used *DNAmAge* and *Ms* within 2-kb windows surrounding 353 CpG-clocks found to correlate with age to infer the age at death of all ancient AMH and AH individuals (excepting K14, ATP2, Motala and RISE523 due to insufficient read depth) (Horvath 2013). This analysis implicitly assumes that the statistical model relating clock-CpG methylation levels throughout the life of modern individuals (Horvath 2013) is valid for ancient populations showing different lifestyles and health conditions. This remains to be tested, although at least some CpG-clock sites are common to humans and apes (Horvath 2013). We found that a large majority of the individuals investigated were estimated to have lived over 50 years (up to 90 years) (supplementary table S4, Supplementary Material online), which is at minimal optimistic for past populations. Additionally, although *DNAmAge* estimates of BR2 (15 years) and Mota (24 years) matched the morphological expectations of a European child and an Ethiopian young adult, the Amerindian Clovis child was estimated to be 34 years old when he died. Further methodological improvements, for example, restricting analyses to clock-CpGs specifically identified in bones and teeth—two tissues not originally represented in the CpG-clock discovery panel (Horvath 2013), are thus needed before this method could be reliably applied.

Secondly, we investigated five particular regions within the *HOXD* cluster and *TBX15* locus, which were part of the ~2,000 differentially methylated regions (DMRs) identified by Gokhman et al. (2014) between AHs and modern osteoblast cells. Given their role during limb development (Zakany and Duboule 2007), these regions were proposed to potentially be involved in the strong anatomical differences between AHs and AMHs. The AH methylation profiles recovered by epiPALEOMIX are largely consistent with those reported by Gokhman et al. (2014) (fig. 10). However, we found similar methylation profiles for both Neand.Altai and the ancient AMH Ust.Ishim at the *HOXD10* and *HOXD9* putative DMRs, except within the ~500-bp region from the promoter. The same was true at the putative *TBX15* DMR. This illustrates how epiPALEOMIX can be used to refine putative DMRs between AMHs and AHs.

Conclusions

In this article, we have developed epiPALEOMIX, the first computational package dedicated to the reconstruction of genome-wide methylation and nucleosome maps in ancient individuals. We applied epiPALEOMIX to the largest panel of genome-scale sequence data from ancient mammals and have shown that genome-wide epigenetic landscapes can be reconstructed over a considerable geographical and temporal span. For all ancient samples showing significant levels of DNA damage and prepared using USER treatment or *Phusion* amplification, inferred methylation profiles are reminiscent of modern methylomes, and show some level of tissue specificity. Importantly, this was also true (although less

pronounced) for some samples that had not been USER treated or *Phusion* amplified due to the faster post-mortem deamination rate at methylated cytosines. By contrasting occupied and nonoccupied CTCF binding regions, we further demonstrate that aDNA fragmentation patterns are driven by nucleosome protection. Nucleosome protection patterns are, however, not found across all samples investigated and do not appear to be age-dependent.

Our epigenome is increasingly acknowledged for participating in gene regulation, and thereby for shaping phenotypes. It can be affected by inherited genetic variation (Heyn et al. 2013) and, during our early embryonic life, by the environmental conditions experienced by our mother, with possible long-term phenotypic and health consequences (Lim and Brunet 2013). Furthermore, epigenetic differences have been found between populations from different continents (Heyn et al. 2013) or exhibiting different lifestyles (Tobi et al. 2009). The epiPALEOMIX package opens for further investigation of epigenetic changes in evolutionary times, in particular during major environmental and cultural transitions (Orlando and Willerslev 2014).

Methods

Methylation Scores, *Ms*

Regional methylation scores (*Ms*) were calculated following the methodology presented by Pedersen et al. (2014) (fig. 1A). The calculation of *Ms* depends on the molecular tools used to prepare aDNA. In method 1-USER, DNA extracts have been USER treated prior to library construction (Briggs et al. 2010) while in method 3-*Phusion*, aDNA libraries based on ligation at A-overhangs are amplified with the *Phusion* Polymerase (Pedersen et al. 2014), a DNA Polymerase which hampers DNA elongation through UpGs (method 3-*Phusion*). Both types of procedures make it possible to track mCpG epialleles through CpG→TpG conversions, and complementary GpC→GpT conversions (Gokhman et al. 2014; Pedersen et al. 2014; Seguin-Orlando, Hoover, et al. 2015). In the third type of experimental procedure (method 2-Regular), DNA extracts have been directly constructed into aDNA libraries, which have been amplified with a DNA polymerase that can elongate through UpG dinucleotides. Here, *Ms* is expected to reflect both contributions of CpG→UpG and mCpG→TpG conversions, resulting in higher *Ms* levels than for method 1-USER and method 3-*Phusion*. However, as post-mortem deamination is faster at methylated cytosines than unmethylated ones (Dianov et al. 1994; Palkopoulou et al. 2015; Seguin-Orlando, Gamba, et al. 2015), *Ms* can still be expected to reflect methylation levels.

As the type of DNA library preparation is known to affect nucleotide mis-incorporation patterns (Meyer et al. 2012; Seguin-Orlando, Hoover, et al. 2015), we accounted for the type of DNA library prepared when counting CpG→TpG mis-incorporations. For DNA library types prepared from double stranded DNA templates, we only considered the 5' end of each read, representing the first 15 positions sequenced. In contrast, for DNA libraries prepared from single stranded DNA templates (Meyer et al. 2012), the 15 positions

from 5' and 3' were considered (Gokhman et al. 2014). Relevant details about the molecular procedures used for each sample, including USER treatment, DNA library type, DNA polymerase, and number of trimmed bases, are summarized in [supplementary table S5, Supplementary Material](#) online. Additionally, for a subset of USER-treated samples, we discarded bases showing high C→T rates at read termini as USER-treatment shows poor efficacy at read termini (Meyer et al. 2012).

Ms was calculated by summing the number of CpGs and TpGs present in the sequencing reads aligned against the plus strand of the reference genome, at genomic positions where the latter displayed a CpG dinucleotide. For sequencing reads aligning against the minus strand, we tracked the presence of CpGs and CpAs in the sequencing reads at positions where the reference genome showed CpGs. Since post-mortem deamination is random and is not expected to affect all genomic mCpG sites, sites for which more than half the reads exhibit CpG→TpG conversion with a minimum coverage of five were disregarded. Similarly, genomic sites described as polymorphic within dbSNP142 (<http://genome.ucsc.edu/>; last accessed September 1, 2016) were disregarded as these can reflect sites where the ancient individual and the reference genome differed. We acknowledge that, in situations where ancient polymorphic sites are unlikely to be represented within dbSNP (e.g., AH and ancient populations with no obvious present-day surrogates), masking procedures specifically accounting for the phylogenetic distance between ancient and present-day populations would be preferable.

Regional Methylation Levels

We used Ms to investigate methylation levels within genomic regions with known methylation patterns, including promoter regions (defined following Ball et al. (2009) as the region comprising the 400-bp upstream and 1,000-bp downstream of TSSs), mtDNA, introns, exons, splice sites, and CpG islands as well as their neighboring shores (flanking 2,000 bases) and shelves (flanking 2,000–4,000 bases). We followed Ball et al. (2009) to stratify annotated promoters into three groups of increasing CpG density displaying an inverse correlation with their methylation levels (Ball et al. 2009). All genomic coordinates for the above-mentioned regions have been downloaded from UCSC table browser (www.genome.ucsc.edu/; last accessed September 1, 2016), using Ensembl gene build from the hg19/build37 reference assembly for humans. Additionally, we retrieved hg18 genomic coordinates of CTCF binding sites from Fu et al. (2008) and converted these to hg19/build37 using *liftOver* (http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/; last accessed September 1, 2016). Our final set of genomic coordinates included 2 kilobases centered at each of 5,735 CTCF occupied and 5,635 unoccupied CTCF binding sites. We excluded sex chromosomes, which display depth-of-coverage variation between males and females, from all our analyses, except when noted otherwise. CGI coordinates for equids were converted from hg19/build37 using *liftOver* to horses (EquCab2.0) and represented 4.9 megabases spread across 7,685 regions. The gene annotation and intron/exon splice

sites coordinates for horses (EquCab2.0) and Aurochs (btau_umd3.1/bosTau6) was retrieved from (www.genome.ucsc.edu/; last accessed September 1, 2016). For details on promoter categories, gene builds and BED files used, see [supplementary table S6, Supplementary Material](#) online. The expected number of deaminations per overhang was estimated within CTCF regions using δ_s and λ parameters as retrieved from mapDamage2 (Jónsson et al. 2013: 2) ([supplementary table S3, Supplementary Material](#) online) and the following formula ($\delta_s \times ((1/\lambda) - 1)$).

Modern Human Methylomes

To compare ancient and present-day methylomes, we downloaded RRBS methylation data for 101 human tissues from UCSC (see [supplementary table S7, Supplementary Material](#) online) (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeHaibMethylRbbs/>; last accessed September 1, 2016). This represented a total number of ~1.6 millions of CpG sites, for which we calculated Ms in their 1,000 flanking bases (i.e., 2,000 bp centered around each CpG). This window size was selected to reflect the correlation of the methylation levels within neighboring regions in the human genome (Eckhardt et al. 2006). As no hair tissue was represented in the RRBS data, we downloaded from UCSC (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeHaibMethyl450/>; last accessed September 1, 2016) methylation data estimated using the Infinium HumanMethylation450 BeadChip system (450K Illumina array; Sandoval et al. 2011; Slieker et al. 2013) for 22 modern human tissues, including hair. We constructed a linear model to compare ancient and present-day human methylomes. It related modern methylation scores, as measured in RRBS experiments and 450K Illumina arrays, with ancient methylation scores, the number of TpGs and CpGs found in the ancient sequencing data, the variance of TpGs as well as the CpG density of the region. All CpG sites covered with less than 10 reads were disregarded from RRBS sequencing data, as the chance events to observe post-mortem deamination events at such sites are limited.

Nucleosome Calls

Nucleosomes were called using the sliding-window approach described by Pedersen et al. (2014) (fig. 1B). This method provides a nucleosome score, which represents a more robust measure of nucleosome presence than simply the maximal read depth. We therefore used this score when aggregating signal across regions, for example, all TSS regions genome-wide, and also disregarded all called nucleosomes with a negative score.

Insert length, GC-content (%GC) and repeats represent well-known factors affecting read depth variation along the genome (Benjamini and Speed 2012; Jónsson et al. 2014), potentially conflicting with our nucleosome calling procedure. To assess whether nucleosome calls were robust to such biases, we restricted nucleosome calls to uniquely mappable genomic regions, which were defined following the procedure described in Jónsson et al. (2014), where we computed

the sequence mappability within 20-kb genomic blocks using a 41-mers (Derrien et al. 2012) and restricted nucleosome calls to blocks showing mappability uniqueness ≥ 0.9 . We also implemented the model-based %GC correction from Benjamini and Speed (Benjamini and Speed 2012) to read depth at each genomic coordinate prior to nucleosome calling. All nucleosome-calling analyses are based on GC-corrected read-depth and uniquely mappable regions unless otherwise noted.

Phasograms

We followed the procedure originally described by Valouev et al. (2011) to analyze nucleosome phasing in HTS data (fig. 1C). We required a minimal depth-of-coverage of three reads per sequencing start position unless stated otherwise.

Regional Nucleosome Profiles

We evaluated ancient data sets at regions known to show strong nucleosome positioning and phasing in a wide range of cell types. One of such regions was represented by CTCF binding sites (defined as above) and another was located on the human chromosome 12 at positions 34,439,733–34,559,733 (hg19) (Gaffney et al. 2012). We also analyzed promoter regions and the respective gene bodies of their longest open reading frame (supplementary table S6, Supplementary Material online).

Fourier Transforms

Periodicity signals were investigated through spectral density analyses, using the *Spectrum* package in R (<http://www.R-project.org/>; last accessed September 1, 2016). This was applied to all genomic regions that displayed fluctuations in either methylation levels, read depth, or phasogram analyses.

The raw data was normalized over the background using LOESS (span = 0.3) function in R (<https://www.cran.r-project.org/>; last accessed September 1, 2016).

Saqqaq Palaeo-Eskimo Sequence Data

We extracted DNA from 150 mg of hair from a Palaeo-Eskimo human sample of the Saqqaq culture, dated $4,044 \pm 31$ BP, following Rasmussen et al. (2010). MBD enrichment was performed on the whole extract as described by Seguin-Orlando, Gamba, et al. (2015). The captured fraction, enriched for methylated DNA, was eluted using a 1-M KCl solution, purified on a MinElute (QIAGEN) column and eluted in 55 μ l EB buffer. An aliquot of the eluate (32.5 μ l) was treated for 3 h at 37 °C using the 10- μ l USER mix (New England Biolabs, #M5505) before an Illumina DNA library was constructed using the procedure from Meyer and Kircher (Meyer and Kircher 2010) with the modifications from Seguin-Orlando, Gamba, et al. (2015). The library was split in two halves, each of which was independently amplified for 15 cycles, using 5 units of AmpliTaq Gold (Life Technologies) and 6-bp indexed primers. After MinElute-purification, both were subjected to a second round of amplification (6 cycles), using four parallel reactions each. Final products were purified on MinElute columns, quantified on the 2100 Bioanalyzer (Agilent) High-Sensitivity Assay, pooled with other libraries and sequenced

on the Illumina HiSeq2500, 75 Paired-End Rapid Mode, at the Danish National High-Throughput DNA sequencing centre. High-quality reads aligning uniquely against the human reference genome hg19 were identified using PALEOMIX (Schubert et al. 2014), disabling seeding in BWA 0.5.9-r26-dev (Li et al. 2009), and requiring a minimum mapping quality of 25. The sequence data are available for download at the European Nucleotide Archive (PRJEB15186).

Supplementary Material

Supplementary figures S1–S13 and tables S1–S7 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

This work was supported by the Danish Council for Independent Research, Natural Sciences (4002-00152B); the Danish National Research Foundation (DNRF94); the Villum Fonden (miGENEPI); and Initiative d'Excellence Chaires d'attractivité, Université de Toulouse (OURASI). L.O. conceived the study; K.H. and L.O. designed the study; A.S.O. performed molecular experiments; E.W. and L.O. provided material and reagents; K.H. performed all analyses, with input from L.O.; K.H. programmed the epiPALEOMIX pipeline, with input from M.S.; L.O. wrote the text, with significant input from K.H., and all other coauthors.

References

- Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, et al. 2015. Population genomics of bronze age Eurasia. *Nature* 522:167–172.
- Ball MP, Li JB, Gao Y, Lee J-H, LeProust EM, Park I-H, Xie B, Daley GQ, Church GM. 2009. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* 27:361–368.
- Benjamini Y, Speed TP. 2012. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res* 40:e72.
- Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, Meyer M, Krause J, Ronan MT, Lachmann M, et al. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* 104:14616–14621.
- Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Pääbo S. 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res* 38:e87.
- Cassidy LM, Martiniano R, Murphy EM, Teasdale MD, Mallory J, Hartwell B, Bradley DG. 2016. Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proc Natl Acad Sci U S A* 113:368–373.
- Dabney J, Knapp M, Glocke I, Gansauge M-T, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga J-L, et al. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A* 110:15758–15763.
- Dabney J, Meyer M. 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTechniques* 52:87–94.
- Deaton AM, Bird A. 2011. CpG islands and the regulation of transcription. *Genes Dev* 25:1010–1022.
- Delsuc F, Gasse B, Sire J-Y. 2015. Evolutionary analysis of selective constraints identifies ameloblastin (AMBN) as a potential candidate for amelogenesis imperfecta. *BMC Evol Biol* 15:148.

- Der Sarkissian C, Ermini L, Schubert M, Yang MA, Librado P, Fumagalli M, Jónsson H, Bar-Gal GK, Albrechtsen A, Vieira FG, et al. 2015. Evolutionary genomics and conservation of the endangered Przewalski's horse. *Curr Biol*. 25:2577–2583.
- Derrien T, Estellé J, Marco Sola S, Knowles DG, Raineri E, Guigó R, Ribeca P. 2012. Fast computation and applications of genome mappability. *PLoS One* 7:e30377.
- Dianov G, Sedgwick B, Daly G, Olsson M, Lovett S, Lindahl T. 1994. Release of 5'-terminal deoxyribose-phosphate residues from incised abasic sites in DNA by the *Escherichia coli* RecJ protein. *Nucleic Acids Res*. 22:993–998.
- Dong Z, Saikumar P, Weinberg JM, Venkatachalam MA. 1997. Internucleosomal DNA cleavage triggered by plasma membrane damage during necrotic cell death. Involvement of serine but not cysteine proteases. *Am J Pathol*. 151:1205–1213.
- Eckhardt F, Lewin J, Cortese R, Rakyán VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, et al. 2006. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet*. 38:1378–1385.
- Fagny M, Patin E, MacIsaac JL, Rotival M, Flutre T, Jones MJ, Siddle KJ, Quach H, Harmant C, McEwen LM, et al. 2015. The epigenomic landscape of African rainforest hunter-gatherers and farmers. *Nat Commun*. 6:10047.
- Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PLF, Aximu-Petri A, Prüfer K, de Filippo C, et al. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514:445–449.
- Fu Q, Mittnik A, Johnson PLF, Bos K, Lari M, Bollongino R, Sun C, Giemsch L, Schmitz R, Burger J, et al. 2013. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol*. 23:553–559.
- Fu Y, Sinha M, Peterson CL, Weng Z. 2008. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet*. 4:e1000138.
- Gaffney DJ, McVicker G, Pai AA, Fondufe-Mittendorf YN, Lewellen N, Michelini K, Widom J, Gilad Y, Pritchard JK. 2012. Controls of nucleosome positioning in the human genome. *PLoS Genet*. 8:e1003036.
- Gallego Llorente M, Jones ER, Eriksson A, Siska V, Arthur KW, Arthur JW, Curtis MC, Stock JT, Coltorti M, Pieruccini P, et al. 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* 350:820–822.
- Gamba C, Hanghøj K, Gaunitz C, Alfarhan AH, Alquraishi SA, Al-Rasheid KAS, Bradley DG, Orlando L. 2016. Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Mol Ecol Resour*. 16:459–469.
- Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, Domboróczki L, Klovári I, Pap I, Anders A, et al. 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nat Commun*. 5:5257.
- Gansauge M-T, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc*. 8:737–748.
- Gansauge M-T, Meyer M. 2014. Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res*. 24:1543–1549.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 5:R80.
- Gokhman D, Lavi E, Prüfer K, Fraga MF, Riancho JA, Kelso J, Pääbo S, Meshorer E, Carmel L. 2014. Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science* 344:523–527.
- Gokhman D, Meshorer E, Carmel L. 2016. Epigenetics: it's getting old. Past meets future in Paleoeugenetics. *Trends Ecol Evol*. 31:290–300.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328:710–722.
- Günther T, Valdiosera C, Malmström H, Ureña I, Rodríguez-Varela R, Sverrisdóttir ÓO, Daskalaki EA, Skoglund P, Naidoo T, Svensson EM, et al. 2015. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc Natl Acad Sci U S A*. 112:11917–11922.
- Hesson LB, Sloane MA, Wong JW, Nunez AC, Srivastava S, Ng B, Hawkins NJ, Bourke MJ, Ward RL. 2014. Altered promoter nucleosome positioning is an early event in gene silencing. *Epigenetics* 9:1422–1430.
- Heyn H, Moran S, Hernandez-Herraez I, Sayols S, Gomez A, Sandoval J, Monk D, Hata K, Marques-Bonet T, Wang L, et al. 2013. DNA methylation contributes to natural human variation. *Genome Res*. 23:1363–1372.
- Hoffmann M, Olson K, Cavender A, Pasqualini R, Gaikwad J, D'Souza RN. 2001. Gene expression in a pure population of odontoblasts isolated by laser-capture microdissection. *J Dent Res*. 80:1963–1967.
- Hofmanová Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Díez-del-Molino D, Dorp L, van López S, Kousathanas A, Link V, et al. 2016. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc Natl Acad Sci*. 113:6886–6891.
- Horvath S. 2013. DNA methylation age of human tissues and cell types. *Genome Biol*. 14:R115.
- Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, et al. 2015. Orchestrating high-throughput genomic analysis with bioconductor. *Nat Methods* 12:115–121.
- Jiang C, Pugh BF. 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet*. 10:161–172.
- Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R, McLaughlin RL, Gallego Llorente M, Cassidy LM, Gamba C, et al. 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun*. 6:8912.
- Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29:1682–1684.
- Jónsson H, Schubert M, Seguin-Orlando A, Ginolhac A, Petersen L, Fumagalli M, Albrechtsen A, Petersen B, Korneliusen TS, Vilstrup JT, et al. 2014. Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc Natl Acad Sci U S A*. 111:18655–18660.
- Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. 2012. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res*. 22:2497–2506.
- Kistler L, Smith O, Ware R, Momber G, Bates R, Garwood P, Fitch S, Pallen M, Gaffney V, Allaby R. 2015. Thermal age, cytosine deamination and the veracity of 8,000 year old wheat DNA from sediments. *bioRxiv* 032060. Available at: <http://biorxiv.org/content/early/2015/11/18/032060>, last accessed September 1, 2016.
- Krause J, Briggs AW, Kircher M, Maricic T, Zwyns N, Derevianko A, Pääbo S. 2010. A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Biol*. 20:231–236.
- Laurent L, Wong E, Li G, Huynh T, Tsigoris A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J, et al. 2010. Dynamic changes in the human methylome during differentiation. *Genome Res*. 20:320–331.
- Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, Fernandes D, Novak M, Gamarra B, Sirak K, et al. 2016. Genomic insights into the origin of farming in the ancient Near East. *Nature* 536:419–424.
- Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513:409–413.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Librado P, Der Sarkissian C, Ermini L, Schubert M, Jónsson H, Albrechtsen A, Fumagalli M, Yang MA, Gamba C, Seguin-Orlando A, et al. 2015. Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proc Natl Acad Sci U S A*. 112:E6889–E6897.
- Lim JP, Brunet A. 2013. Bridging the transgenerational gap with epigenetic memory. *Trends Genet*. 29:176–186.

- Llamos B, Holland ML, Chen K, Cropley JE, Cooper A, Suter CM. 2012. High-resolution analysis of cytosine methylation in ancient DNA. *PLoS One* 7:e30226.
- Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389:251–260.
- Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, et al. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454:766–770.
- Metzker ML. 2010. Sequencing technologies – the next generation. *Nat Rev Genet.* 11:31–46.
- Meyer M, Arsuaga J-L, de Filippo C, Nagel S, Aximu-Petri A, Nickel B, Martínez I, Gracia A, de Castro JMB, Carbonell E, et al. 2016. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* 531:504–507.
- Meyer M, Fu Q, Aximu-Petri A, Glocke I, Nickel B, Arsuaga J-L, Martínez I, Gracia A, de Castro JMB, Carbonell E, et al. 2014. A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* 505:403–406.
- Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* 2010:db.prot5448.
- Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338:222–226.
- Niyibizi C, Eyre DR. 1994. Structural characteristics of cross-linking sites in type V collagen of bone. Chain specificities and heterotypic links to type I collagen. *Eur J Biochem.* 224:943–950.
- Olalde I, Allentoft ME, Sánchez-Quinto F, Santpere G, Chiang CWK, DeGiorgio M, Prado-Martinez J, Rodríguez JA, Rasmussen S, Quilez J, et al. 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* 507:225–228.
- Orlando L, Gilbert MTP, Willerslev E. 2015. Reconstructing ancient genomes and epigenomes. *Nat Rev Genet.* 16:395–408.
- Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, et al. 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499:74–78.
- Orlando L, Willerslev E. 2014. Evolution. An epigenetic window into the past?. *Science* 345:511–512.
- Ozsolak F, Song JS, Liu XS, Fisher DE. 2007. High-throughput mapping of the chromatin structure of human promoters. *Nat Biotechnol.* 25:244–248.
- Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, Li H, Omrak A, Vartanyan S, Poinar H, Götherström A, et al. 2015. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr Biol.* 25:1395–1400.
- Park SDE, Magee DA, McGettigan PA, Teasdale MD, Edwards CJ, Lohan AJ, Murphy A, Braud M, Donoghue MT, Liu Y, et al. 2015. Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle. *Genome Biol.* 16:234.
- Pedersen JS, Valen E, Velazquez AMV, Parker BJ, Rasmussen M, Lindgreen S, Lilje B, Tobin DJ, Kelly TK, Vang S, et al. 2014. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res.* 24:454–466.
- Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505:43–49.
- Racimo F, Renaud G, Slatkin M. 2016. Joint estimation of contamination, error and demography for nuclear DNA from ancient humans. *PLoS Genet.* 12:e1005972.
- Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW, Jr, Rasmussen S, Moltke I, Albrechtsen A, Doyle SM, et al. 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506:225–229.
- Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, et al. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463:757–762.
- Rebello AP, Williams SL, Moraes CT. 2009. In vivo methylation of mtDNA reveals the dynamics of protein-mtDNA interactions. *Nucleic Acids Res.* 37:6701–6715.
- Renaud G, Slon V, Duggan AT, Kelso J. 2015. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* 16:224.
- Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M. 2011. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 6:692–702.
- Schubert M, Ermini L, Der Sarkissian C, Jónsson H, Ginolhac A, Schaefer R, Martin MD, Fernández R, Kircher M, McCue M, et al. 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc.* 9:1056–1082.
- Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, Ginolhac A, Albrechtsen A, Dupanloup I, Foucal A, Petersen B, et al. 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc Natl Acad Sci U S A.* 111:E5661–E5669.
- Seguin-Orlando A, Gamba C, Der Sarkissian C, Ermini L, Louvel G, Boulygina E, Sokolov A, Nédoluzhko A, Lorenzen ED, Lopez P, et al. 2015. Pros and cons of methylation-based enrichment methods for ancient DNA. *Sci Rep.* 5:11826.
- Seguin-Orlando A, Hoover CA, Vasiliev SK, Ovodov ND, Shapiro B, Cooper A, Rubin EM, Willerslev E, Orlando L. 2015. Amplification of TruSeq ancient DNA libraries with AccuPrime Pfx: consequences on nucleotide misincorporation and methylation patterns. *Sci Technol Archaeol Res.* 1:1–9.
- Seguin-Orlando A, Korneliusson TS, Sikora M, Malaspina A-S, Manica A, Moltke I, Albrechtsen A, Ko A, Margaryan A, Moiseyev V, et al. 2014. Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* 346:1113–1118.
- Slieker RC, Bos SD, Goeman JJ, Bovée JV, Talens RP, van der Breggen R, Suchiman HED, Lameijer E-W, Putter H, van den Akker EB, et al. 2013. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics Chromatin.* 6:26.
- Smith RWA, Monroe C, Bolnick DA. 2015. Detection of Cytosine methylation in ancient DNA from five native American populations using bisulfite sequencing. *PLoS One* 10:e0125344.
- Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. 2016. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell* 164:57–68.
- Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, Stein AD, Slagboom PE, Heijmans BT. 2009. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet.* 18:4046–4053.
- Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. 2011. Determinants of nucleosome organization in primary human cells. *Nature* 474:516–520.
- Wang J-P, Fondufe-Mittendorf Y, Xi L, Tsai G-F, Segal E, Widom J. 2008. Preferentially quantized linker DNA lengths in *Saccharomyces cerevisiae*. *PLoS Comput Biol.* 4:e1000175.
- Zakany J, Duboule D. 2007. The role of Hox genes during vertebrate limb development. *Curr Opin Genet Dev.* 17:359–366.
- Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlhauser O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, et al. 2013. Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500:477–481.